Introduction to epidemiology

Course: 'Epidemiological methods in medical research'

Katrine Strandberg-Larsen
Section of Epidemiology

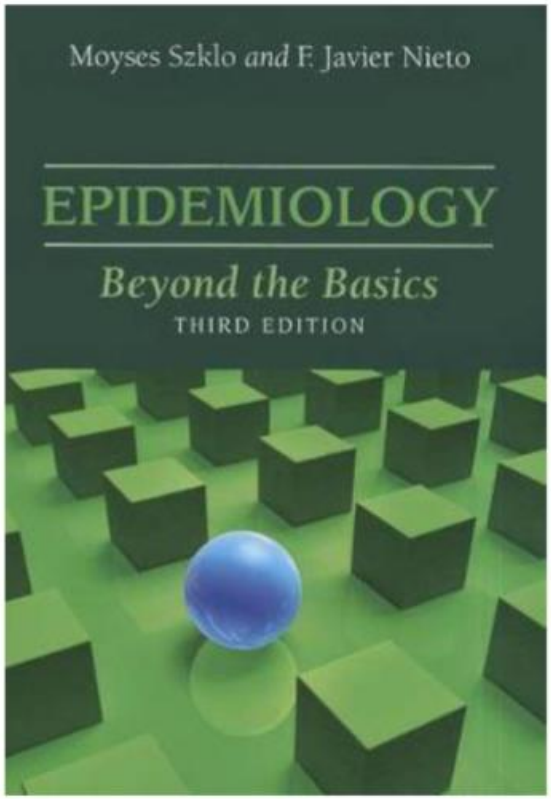
UNIVERSITY OF COPENHAGEN

See you again

- January 30: Case-control studies
  - If you are working on a study with this design, and interested in feedback then approach me in the break
- March 20: Reproducibility, replication & good practice in Epidemiology + Mini seminar

Learning objectives

- Understand the basic principles in epidemiology
- Reflect on knowledge generated using epi methods
- Understand the utility of monitoring health and disease
- Understand the inferential approach how hypothesis are generated and tested
- Understand how causes of diseases are identified
- Understand advantages of using different study designs
- Introduction to the concept of bias, and strategies to identify and reduce bias in epi studies



Presentation

- Name
- Affiliation
- Educational background
- PhD project
- Types of epidemiological designs that you are using/planning to use
- Experience with epidemiological research

What is epidemiology?

Etymology

- Epi = on, with
- Demos = people, population
- Logos = knowledge of
- Epidemiology = The knowledge of what lies on people

Purpose

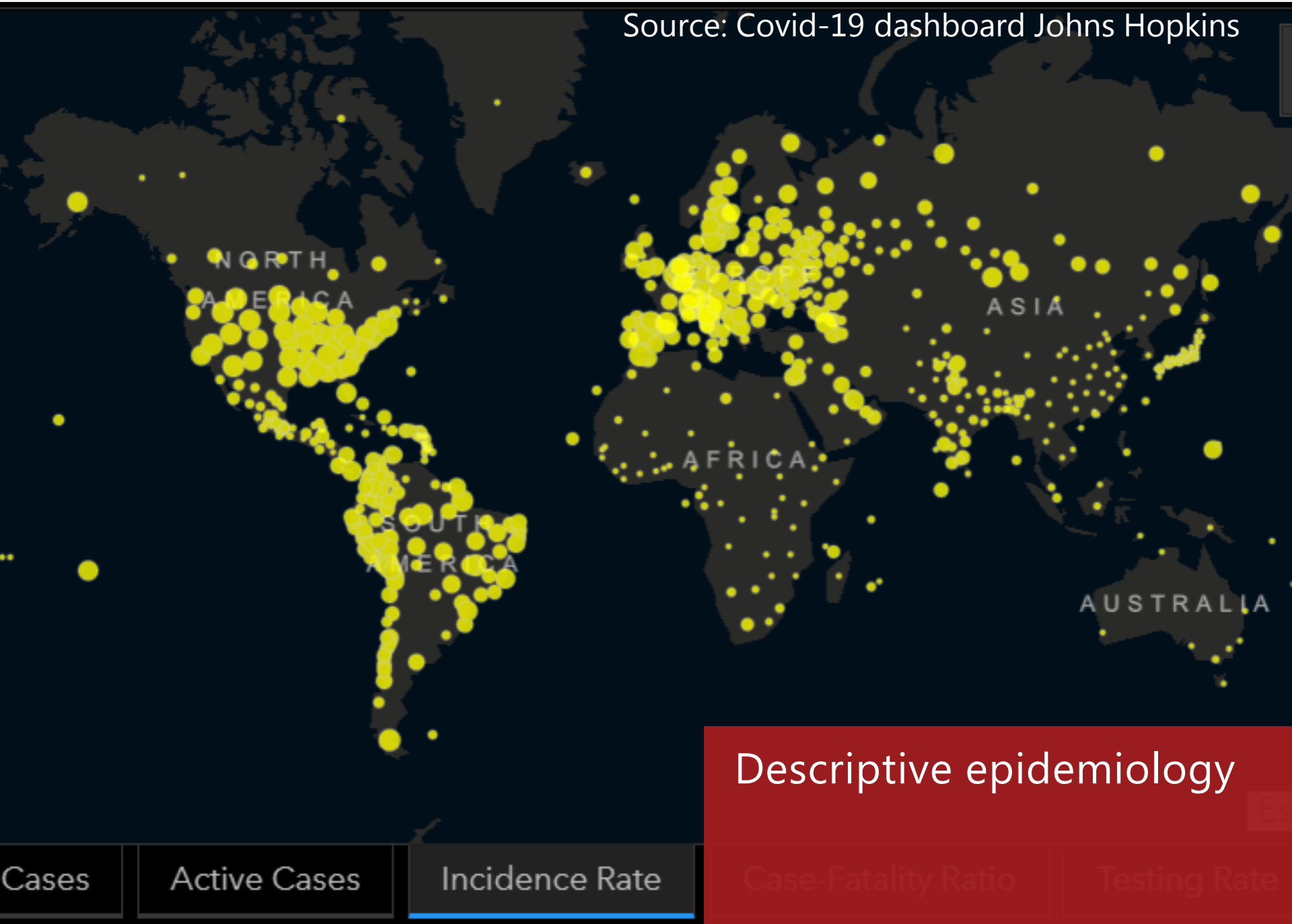
- Monitor health and illness in populations – determine burden of disease
- Identification of causes and risk factors of disease
- Study natural history and prognosis of disease
- Evaluate preventive and therapeutic interventions
- Provide evidence for public health policy and clinical practice

The unit of interest

Arche distinctions in epidemiology

DESCRIPTIVE	ANALYTICAL
The distribution of health-related states or determinants in defined populations	**Public health** The population consists of healthy people and the focus is examining the transition from healthy to disease **Clinical** The population consists of patients and the focus is prognosis, i.e. examining transition from illness to getting healthy again or die

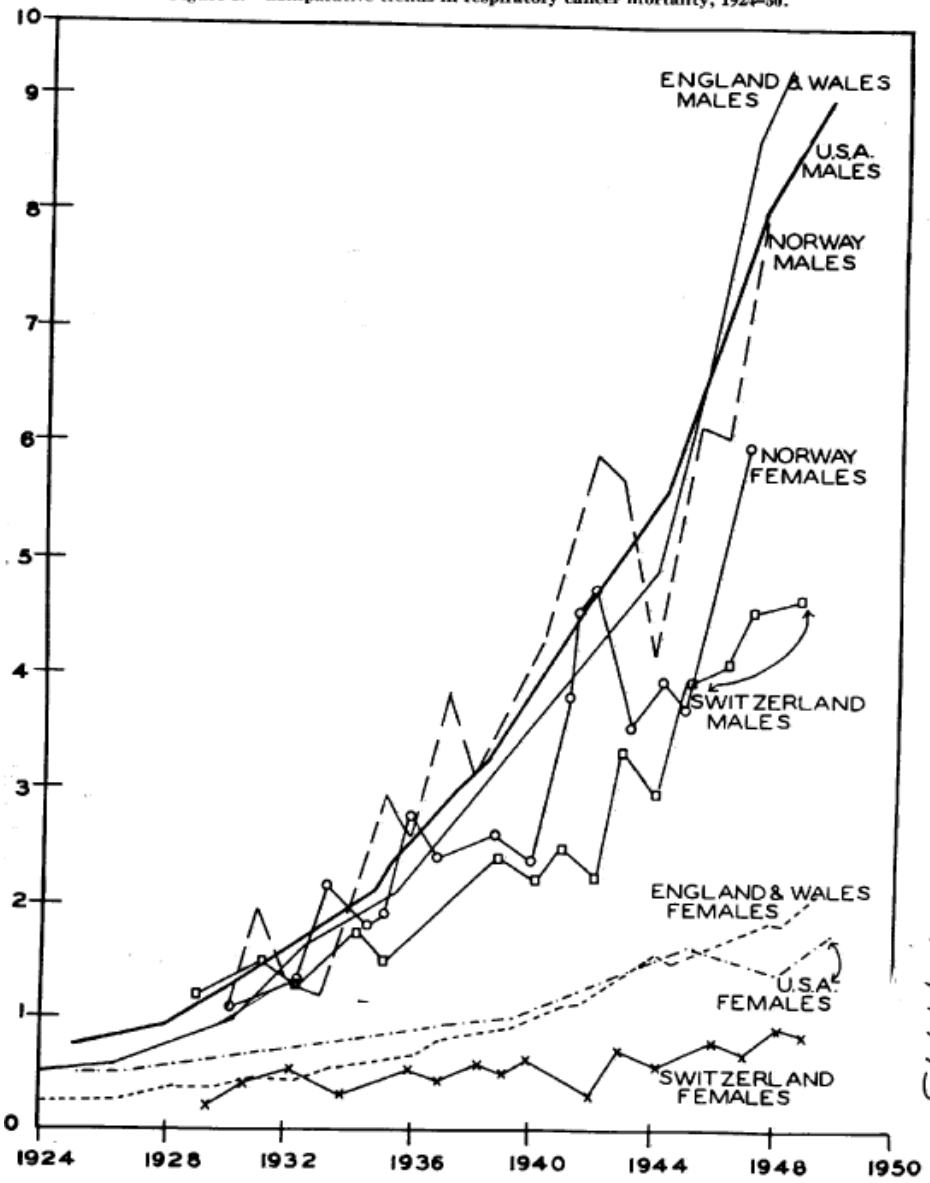
Source: Covid-19 dashboard Johns Hopkins



Descriptive epidemiology

Any pattern?

Figure 1. Comparative trends in respiratory cancer mortality, 1924–50.



01149743

Increased lung cancer mortality

- Steep and continuous increase in lung cancer mortality during the first part of the 20th century in Western countries.

- For example, the lung cancer mortality of English men increased from 1,1 pr. 100.000 in 1901-20 to 10,6 pr. 100.000 in 1936-9.

- In UK, the lung cancer mortality increased 6 times for men and 3 times for women from 1921-30 to 1940-44.

Figure 6. Trends in selected environmental factors, United States, 1900–53 (Hammond).

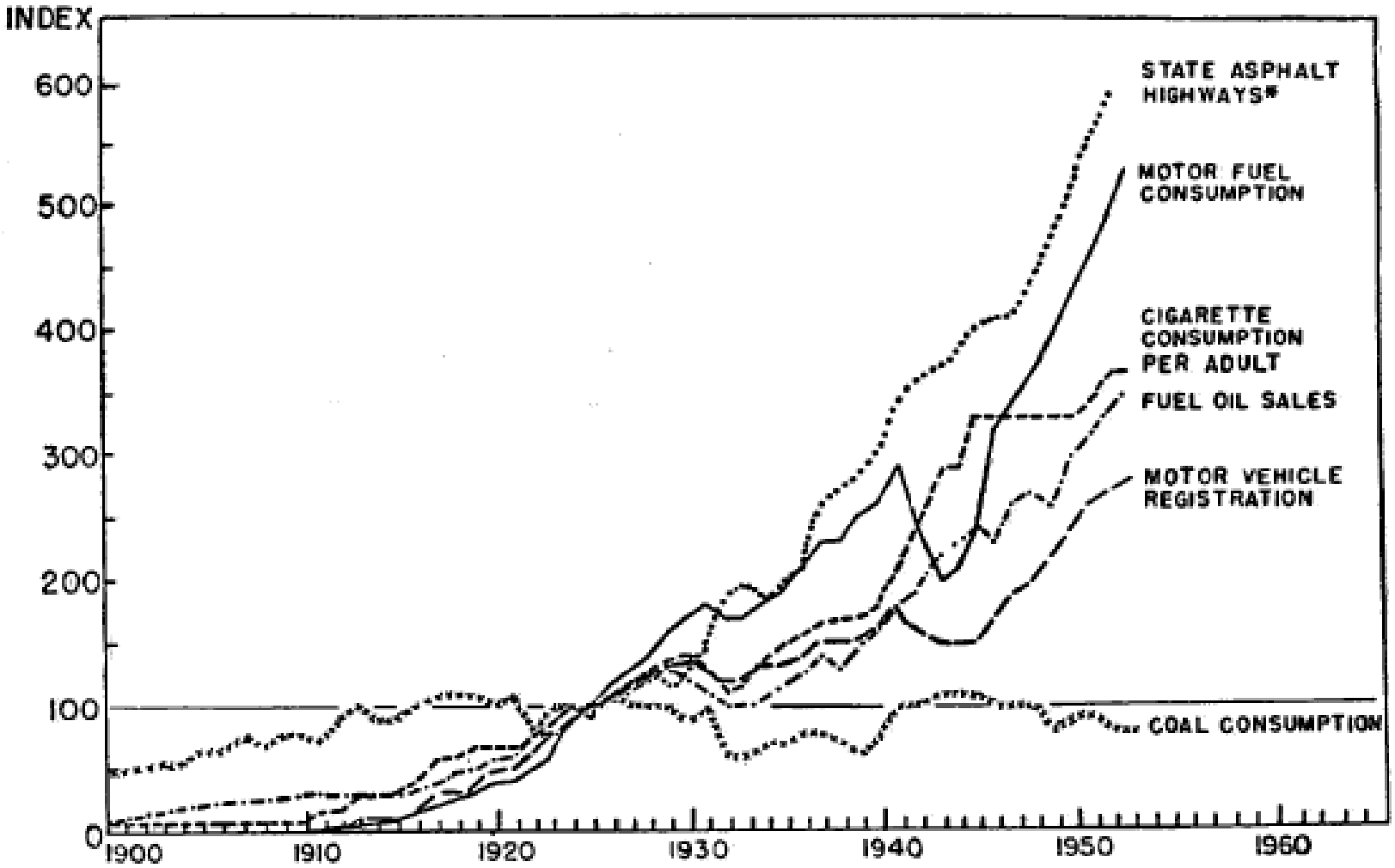
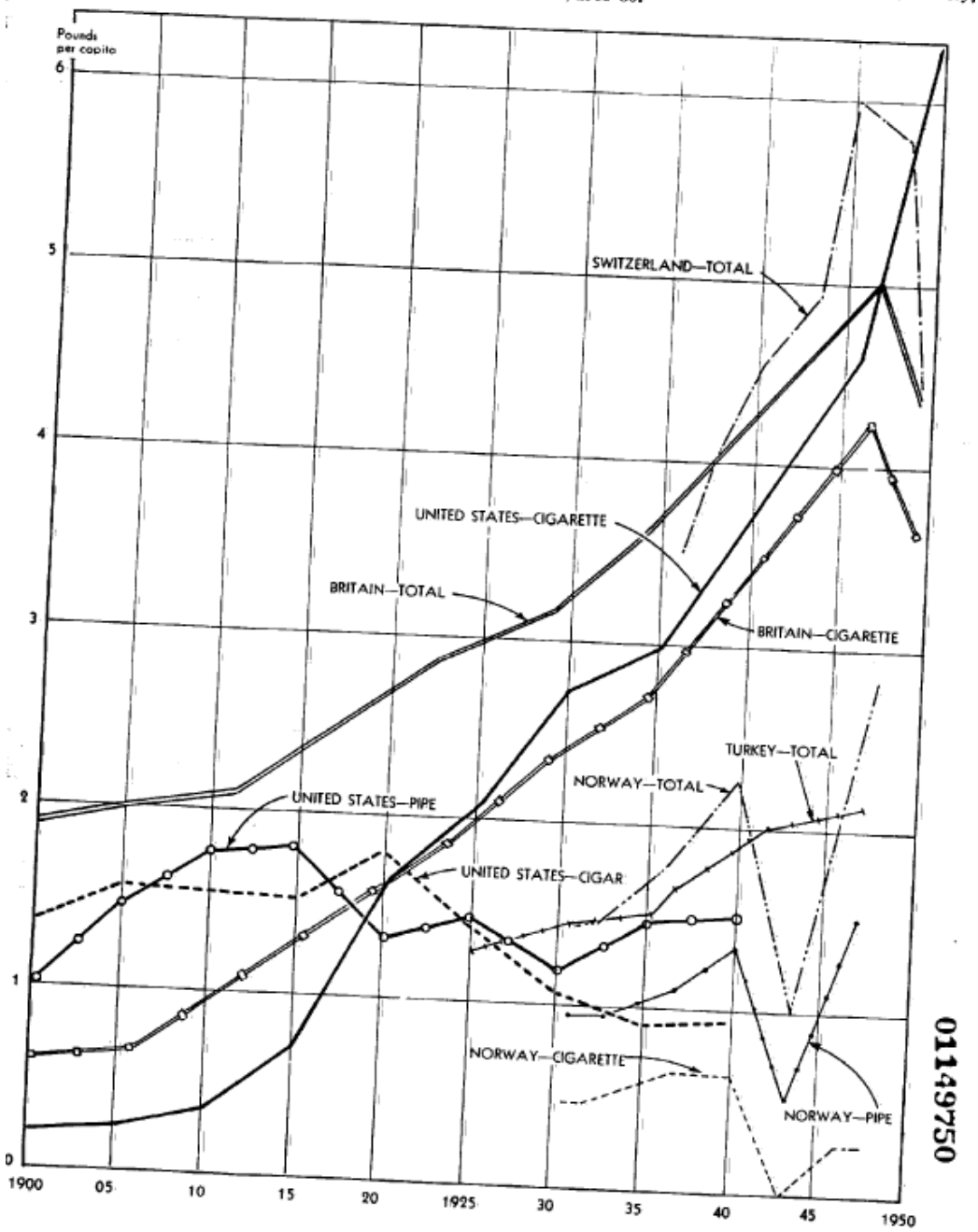


Figure 5. Annual tobacco consumption in pounds per capita, Great Britain, Norway, Switzerland, Turkey, and the United States, 1900–50.

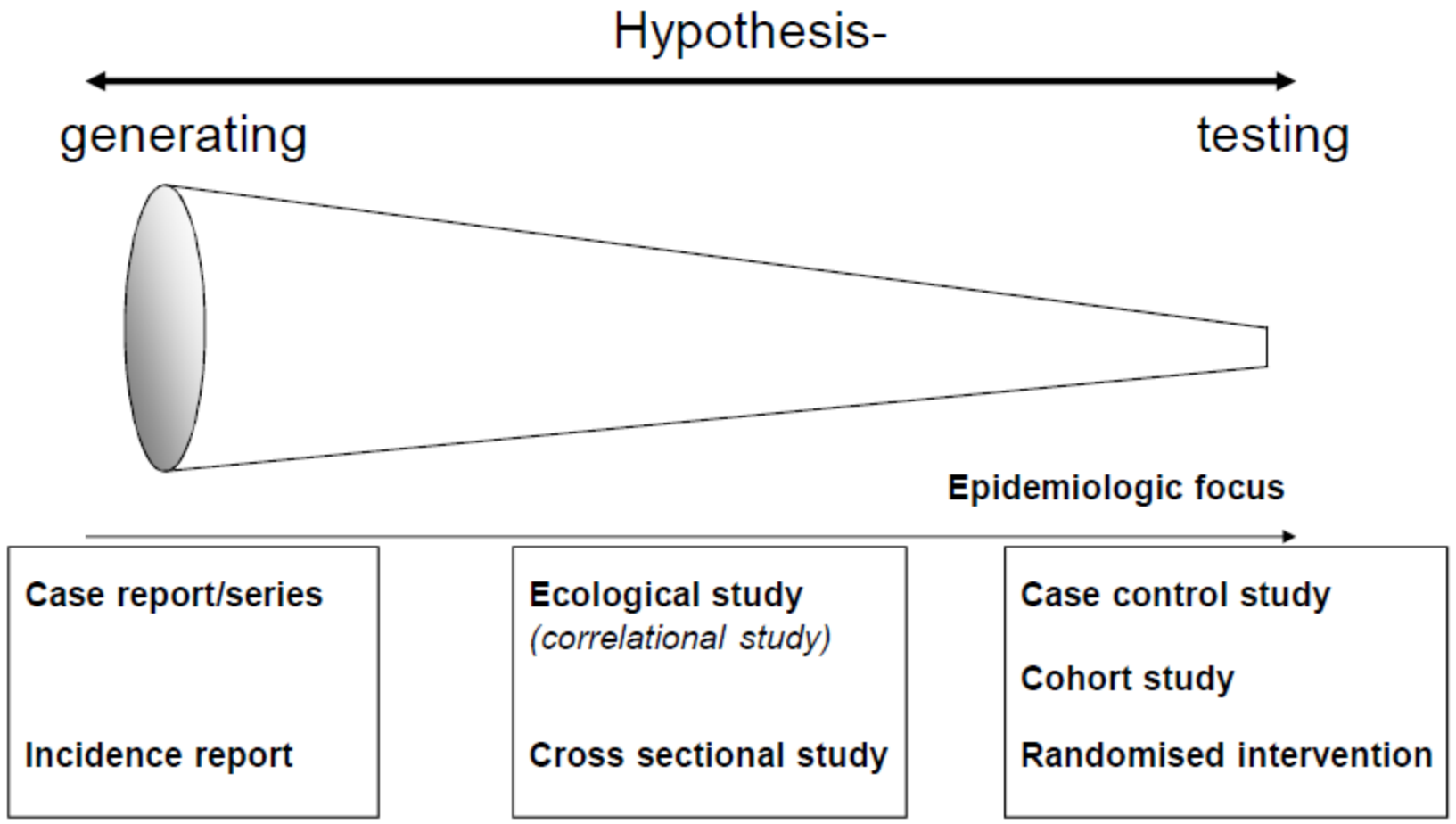


The role of epidemiology

Is to assembling a diverse body of facts from different sources into a coherent explanation.

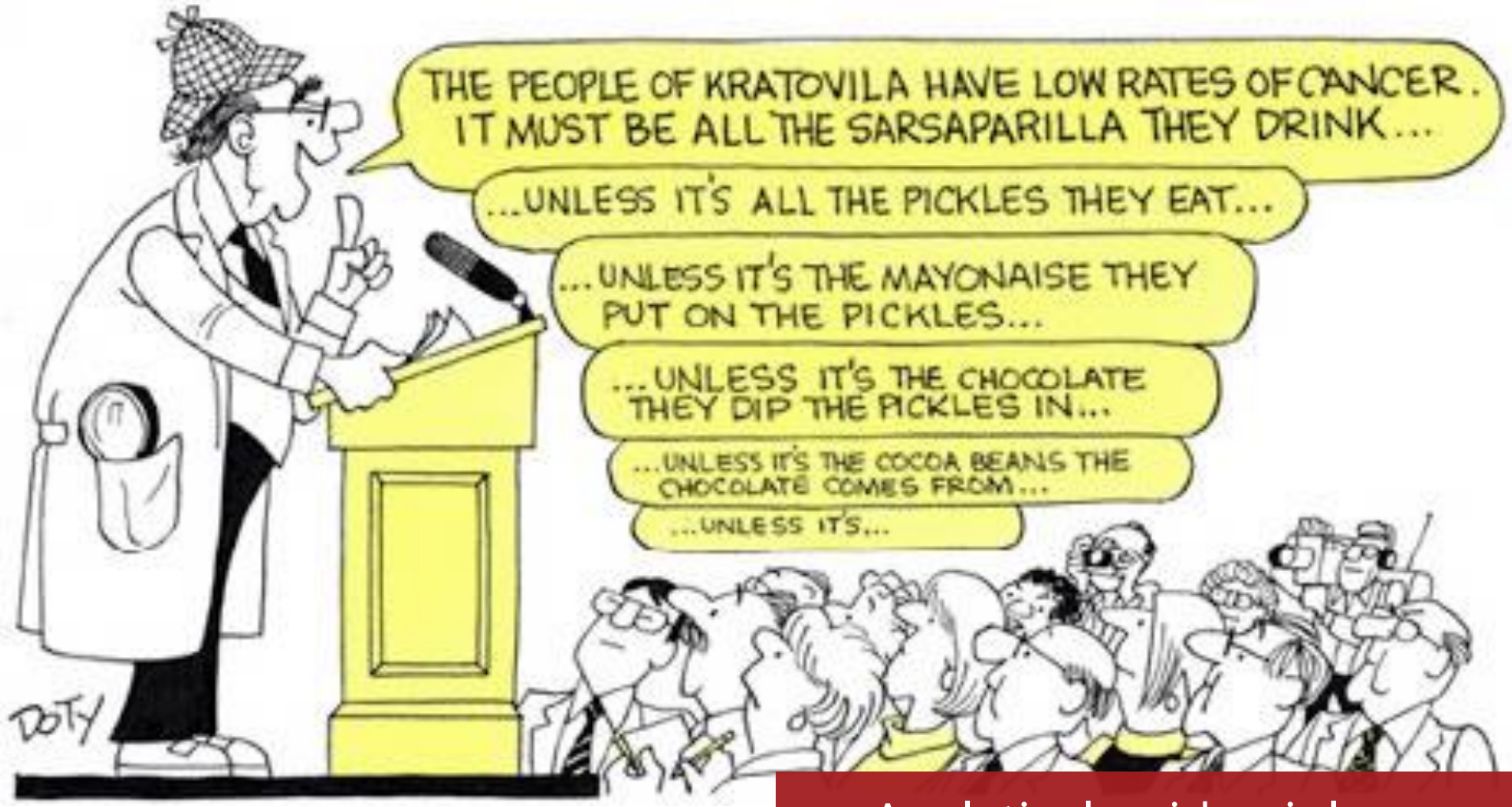
> *"Epidemiology is something more than the total of its established facts. It includes their orderly arrangement into chains of inference which extend more or less beyond the bonds of direct observation"*
>
> American epidemiologist Wade Hampton Frost

Assembling and systemizing evidence



The evidence hierarchy

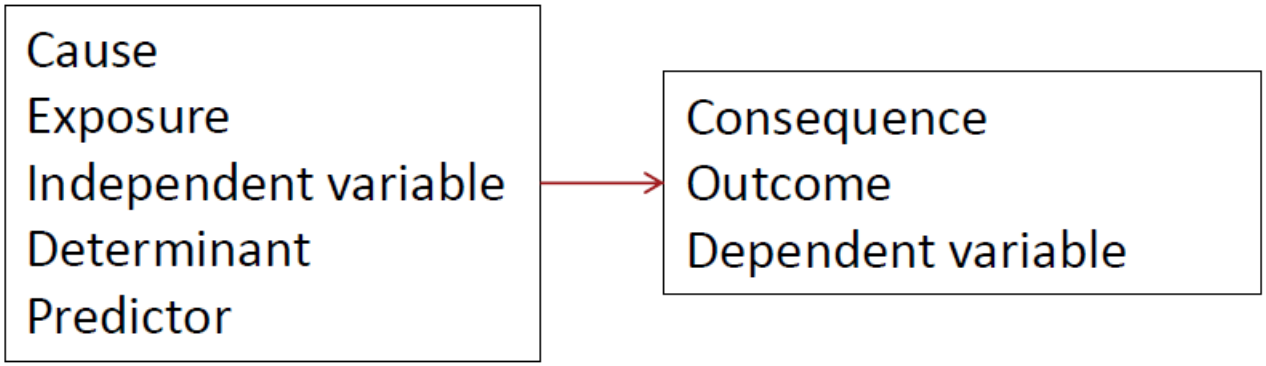
http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/



Analytical epidemiology

What causes disease?

Same, same, but different



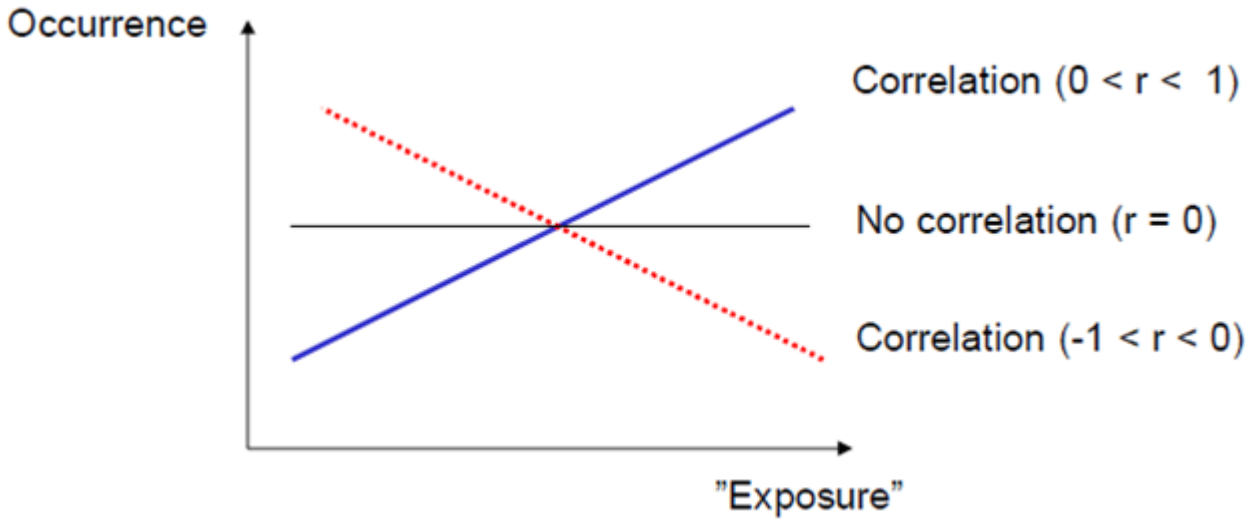
An association is not the same as a causal relationship!!!

This is where epidemiology becomes more than calculating

Ecological/correlation studies

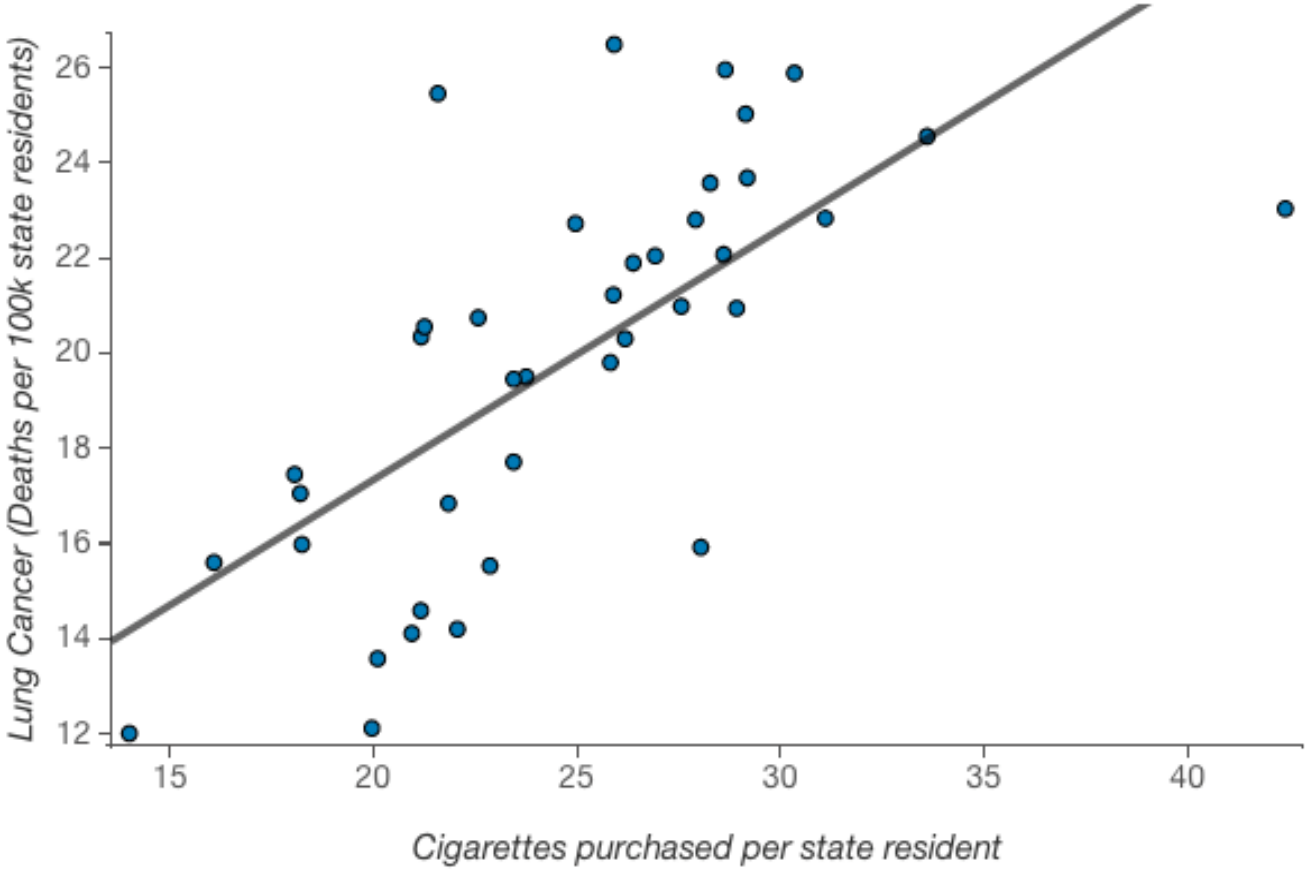
- Make use of available aggregated data, i.e. group level

- Plot outcome frequencies against exposure frequencies in e.g. countries or states



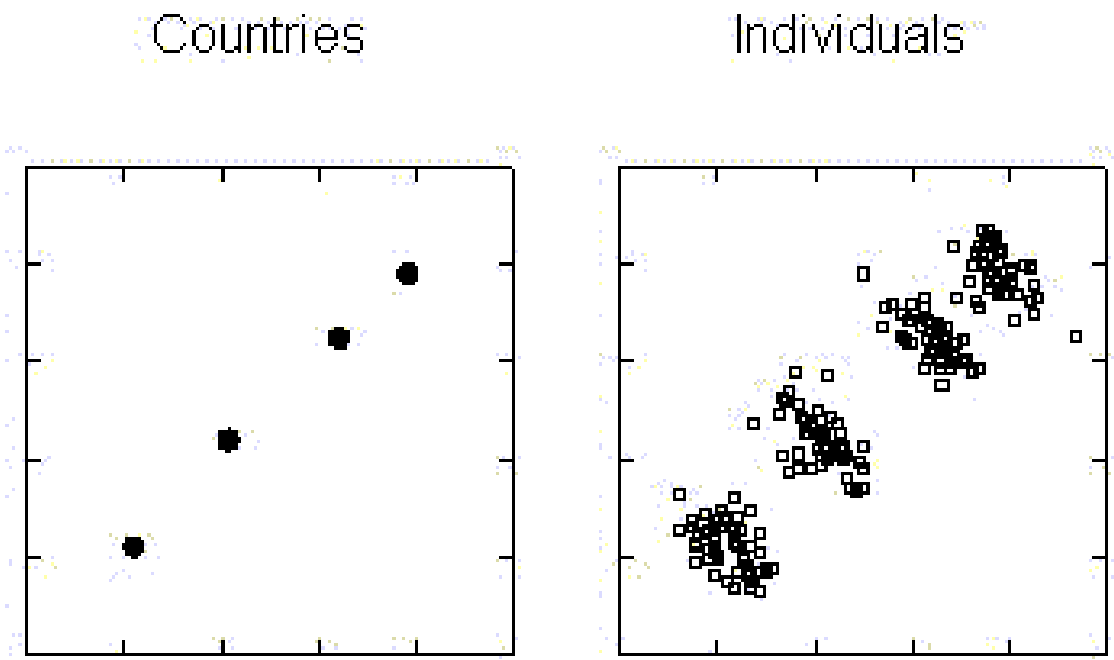
1968 data on cigarette consumption and lung cancer

- What to infer from these data – does this supports that smoking causes lung cancer?

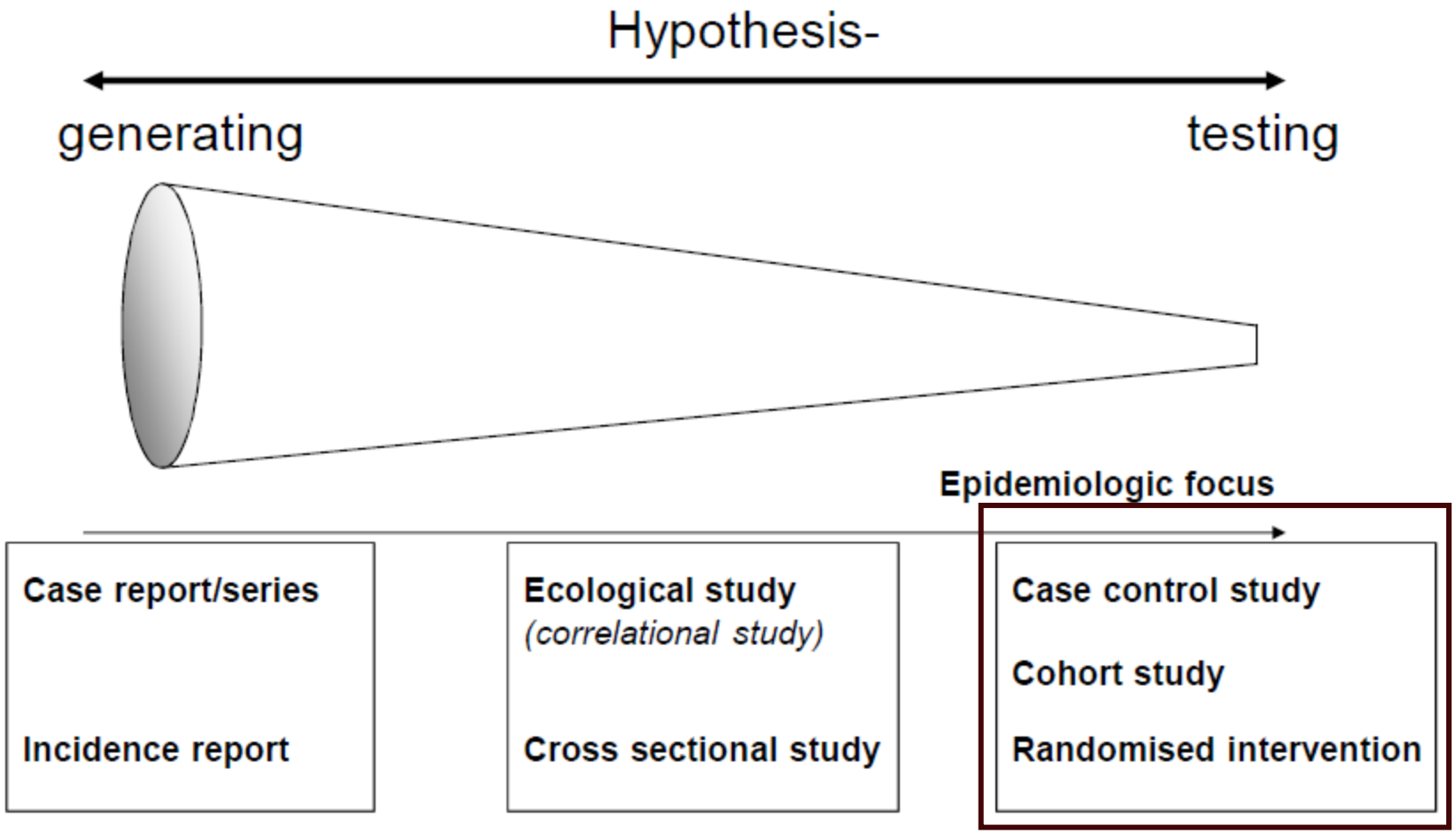


The ecologic fallacy

Bias that may occur because an association observed between variables on an aggregate level does not necessarily represent the association that exists at an individual level



Assembling and systemizing evidence



Does smoking causes lung cancer?

Patients with lung cancer

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

In England and Wales the phenomenal increase in the number of deaths attributed to cancer of the lung provides one of the most striking changes in the pattern of mortality recorded by the Registrar-General. For example, in the quarter of a century between 1922 and 1947 the annual number of deaths recorded increased from 612 to 9,287, or roughly fifteenfold. This remarkable increase is, of course, out of all proportion to the increase of population—both in total and, particularly, in its older age groups. Stocks (1947), using standardized death rates to allow for these population changes, shows the following trend: rate per 100,000 in 1901–20, males 1.1, females 0.7; rate per 100,000 in 1936–9, males 10.6, females 2.5. The rise seems whole explanation, although no one would deny that it may well have been contributory. As a corollary, it is right and proper to seek for other causes.

Possible Causes of the Increase

Two main causes have from time to time been put forward: (1) a general atmospheric pollution from the exhaust fumes of cars, from the surface dust of tarred roads, and from gas-works, industrial plants, and coal fires; and (2) the smoking of tobacco. Some characteristics of the former have certainly become more prevalent in the last 50 years, and there is also no doubt that the smoking of cigarettes has greatly increased. Such associated changes

Comparisons of cases and controls

TABLE II.—*Comparison Between Lung-carcinoma Patients and Non-cancer Patients Selected as Controls, With Regard to Sex, Age, Social Class, and Place of Residence*

Age	No. of Lung-carcinoma Patients		No. of Non-cancer Control Patients	
	M	F	M	F
25– ..	2	1	2	1
30– ..	6	0	6	0
35– ..	18	3	18	3
40– ..	36	4	36	4
45– ..	87	10	87	10
50– ..	130	11	130	11
55– ..	145	9	145	9
60– ..	109	9	109	9
65– ..	88	9	89*	9
70–74 ..	28	4	27*	4
All ages	649	60	649	60

Social Class (Registrar-General's Categories. Men Only)	No. of Lung-carcinoma Patients	No. of Non-cancer Patients
I and II ..	77	87
III	388	396
IV and V ..	184	166
All classes ..	649	649
Place of residence		
County of London	330	377
Outer London	203	231
Other county borough ..	23	16
Urban district	95	54
Rural district ..	43	27
Abroad or in Services ..	15	4
Total (M + F) ..	709	709

* One control patient was selected, in error, from the wrong age group.

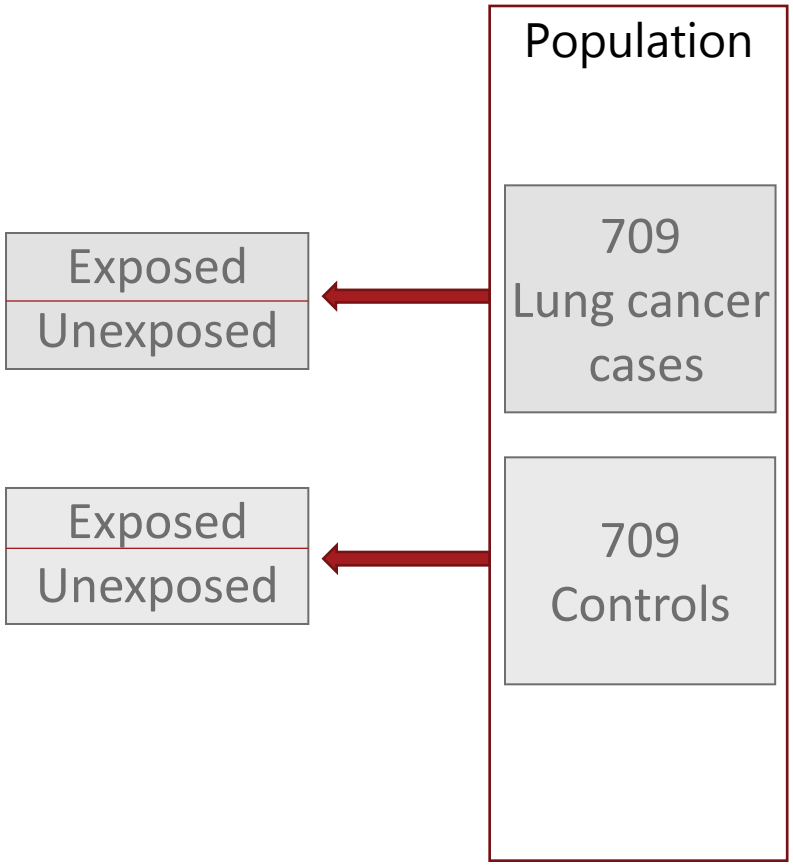
Case-control design

In brief:

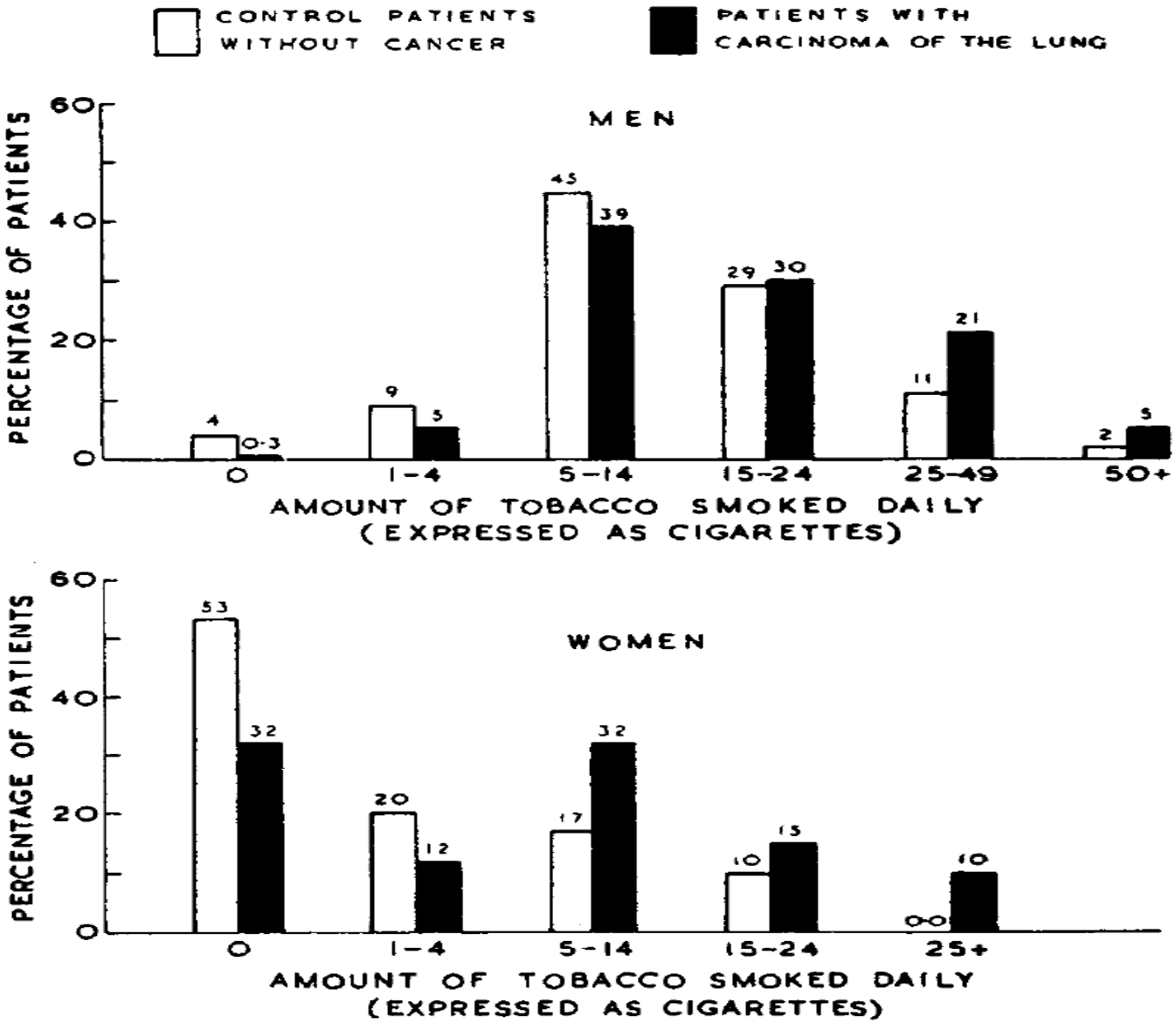
Collect data on exposure on a group of cases and a group of controls

Compare the exposure frequency in cases and controls

The selection of controls must not depend upon exposure



What did they observe?



Conclusion

To summarize, it is not reasonable, in our view, to attribute the results to any special selection of cases or to bias in recording. In other words, it must be concluded that there is a real association between carcinoma of the lung and smoking.

Doll and Hill. BMJ 1950

Tried a new approach

BRITISH MEDICAL JOURNAL

LONDON SATURDAY NOVEMBER 10 1956

LUNG CANCER AND OTHER CAUSES OF DEATH IN RELATION TO SMOKING

A SECOND REPORT ON THE MORTALITY OF BRITISH DOCTORS

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

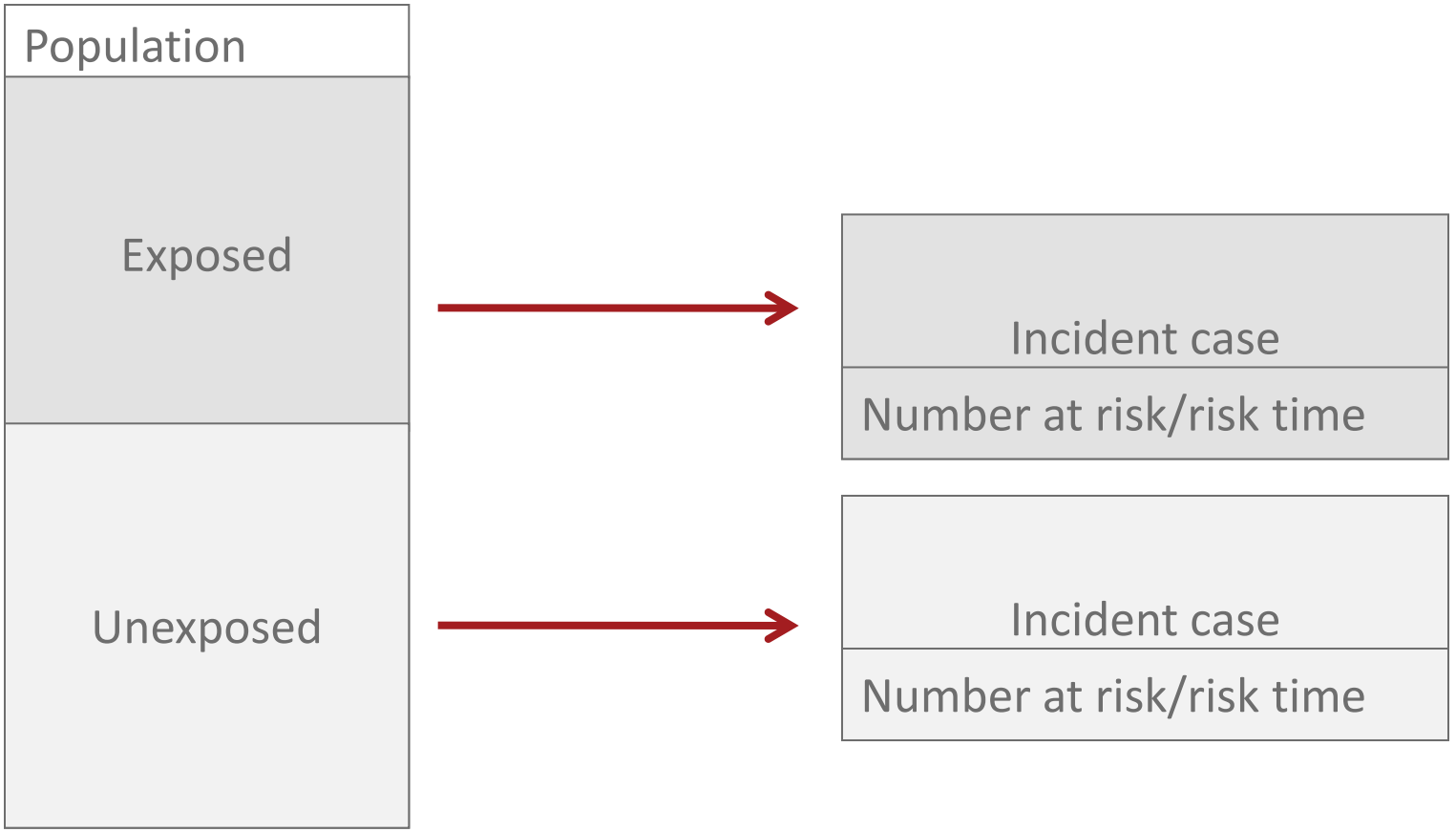
A. BRADFORD HILL, C.B.E., F.R.S.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

On October 31, 1951, we sent a simple questionary to all members of the medical profession in the United Kingdom. In addition to giving their name, address, and age, they were asked to classify themselves into one of three groups—namely, (*a*) whether they were, at that time, smokers of tobacco; (*b*) whether they had smoked but had given up; or (*c*) whether they had never smoked regularly (which we defined as having never smoked as much as one cigarette a day, or its equivalent in pipe tobacco or cigars, for as long as one year). All smokers and ex-smokers were asked additional questions. The previously have been a light smoker or may since then have given up smoking altogether; we shall have continued to count him, or her, as a heavy smoker. If there is a differential death rate with smoking, we must by such errors tend to inflate the mortality among the light smokers and to reduce the mortality among the heavy smokers. In other words, the gradients we present in this paper may be understatements but (apart from sampling errors due to the play of chance) cannot be overstatements.

In 1954 we published a preliminary report on the

Cohort design



What they observed after 5 years?

TABLE V.—*Standardized Death Rates Per Year Per 1,000 Men Aged 35 Years or More, in Relation to the Most Recent Amount Smoked**

Cause of Death	No. of Deaths	Death Rate Among:					
		All Men	Non-smokers	All Smokers	Men Smoking a Daily Average of		
					1–14 g.	15–24 g.	25 g. or More
Lung cancer ..	84†	0·81	0·07	0·90	0·47	0·86	1·66
Other cancer ..	220	2·02	2·04	2·02	2·01	1·56	2·63
Other respiratory diseases ..	126	1·10	0·81	1·13	1·00	1·11	1·41
Coronary thrombosis ..	508	4·78	4·22	4·87	4·64	4·60	5·99
Other causes ..	779	6·79	6·11	6·89	6·82	6·38	7·19
All causes ..	1,714	15·48	13·25	15·78	14·92	14·49	18·84

* That is, at November 1, 1951, for those smoking at that time and at the date of giving up for those who had given up at November 1, 1951.

† The three cases in which lung cancer was recorded as a contributory but not a direct cause of death are included under both lung cancer and the cause to which death was assigned by the Registrar-General.

Doll and Hill. BMJ 1956

What they observed after 10 years?

TABLE 21.—*Standardized Death Rates from Other Diseases*

Cause of Death	No. of Deaths	Death Rate per 1,000					
		All Men	Non-smokers	All Smokers	Cigar-ette Smokers	Mixed Smokers	Pipe or Cigar Smokers
Peptic ulcer ..	54*	0·17	0·03	0·18	0·21	0·16	0·12
Cirrhosis of liver and alcoholism	33	0·10	0·00	0·11	0·12	0·11	0·05
Other digestive diseases ..	87	0·26	0·07	0·28	0·32	0·20	0·25
Genito-urinary diseases† ..	82	0·24	0·33	0·24	0·27	0·21	0·22
Indefinite causes	50	0·15	0·17	0·14	0·13	0·12	0·20
Violence ..	248	0·77	0·94	0·75	0·79	0·68	0·64
All remaining causes of death	150	0·46	0·50	0·47	0·49	0·48	0·45

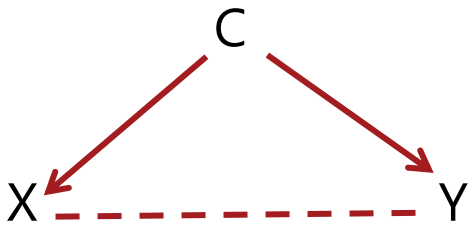
* Including 15 deaths in which peptic ulcer was certified as being associated with the death but not its direct or underlying cause.
† Excluding nephritis.

Doll and Hill. BMJ 1964

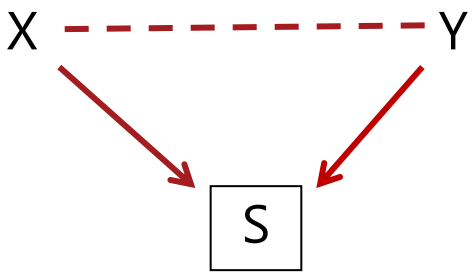
Two variables are associated if…

1. Causal effect

2. Share a common cause



3. Condition on a common effect

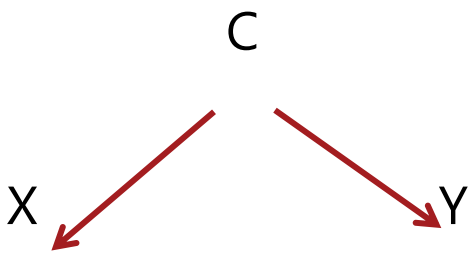


Type of biases

1. Causal effect

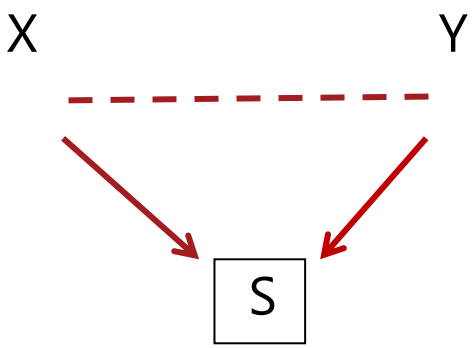
Reverse causation

2. Share a common cause



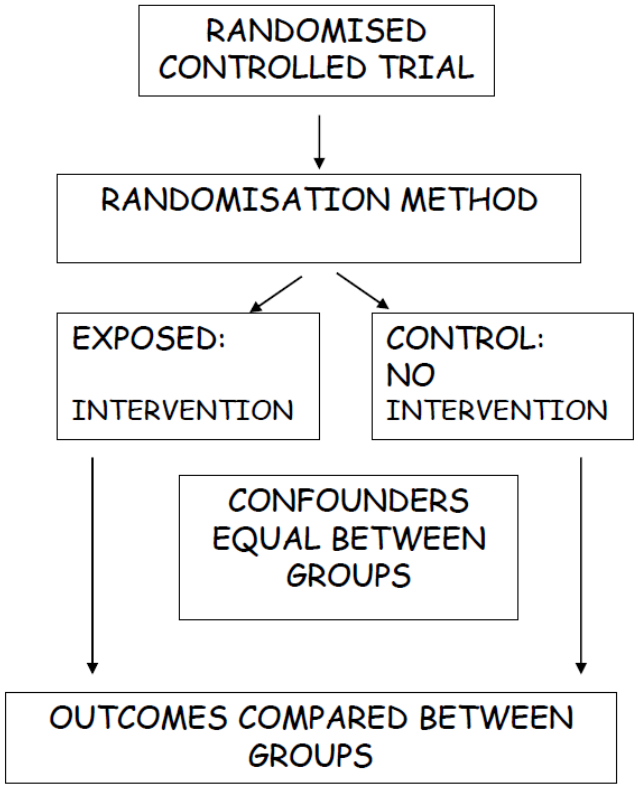
Confounding

3. Condition on a common effect



Collider stratification bias, incl. selection bias

Randomization



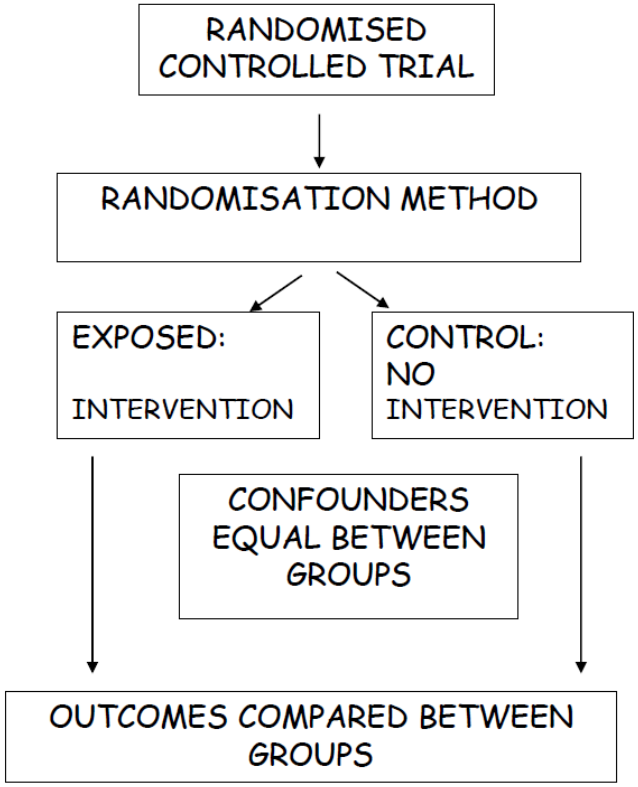
Can we use this design to:

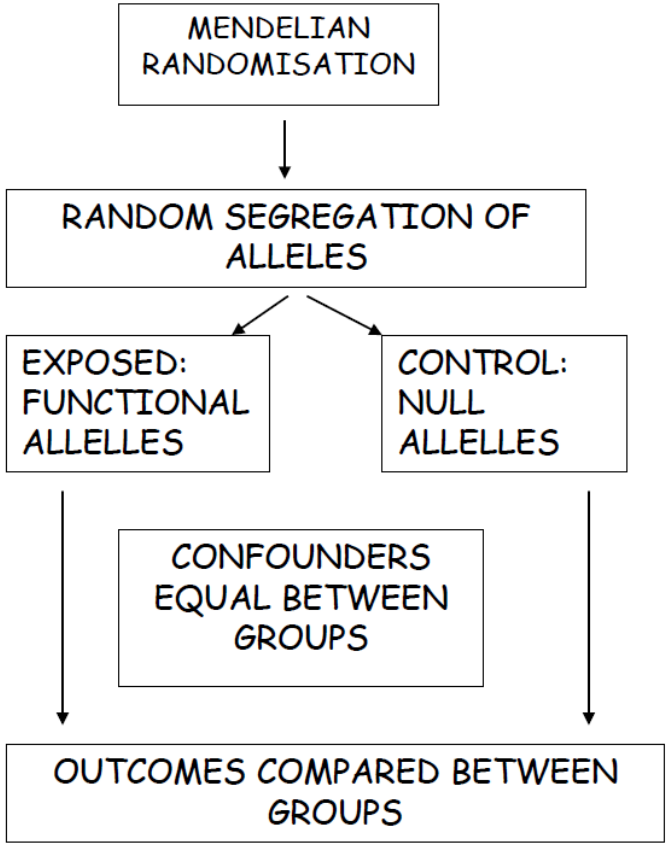
Smoking → Lung cancer

Exercise

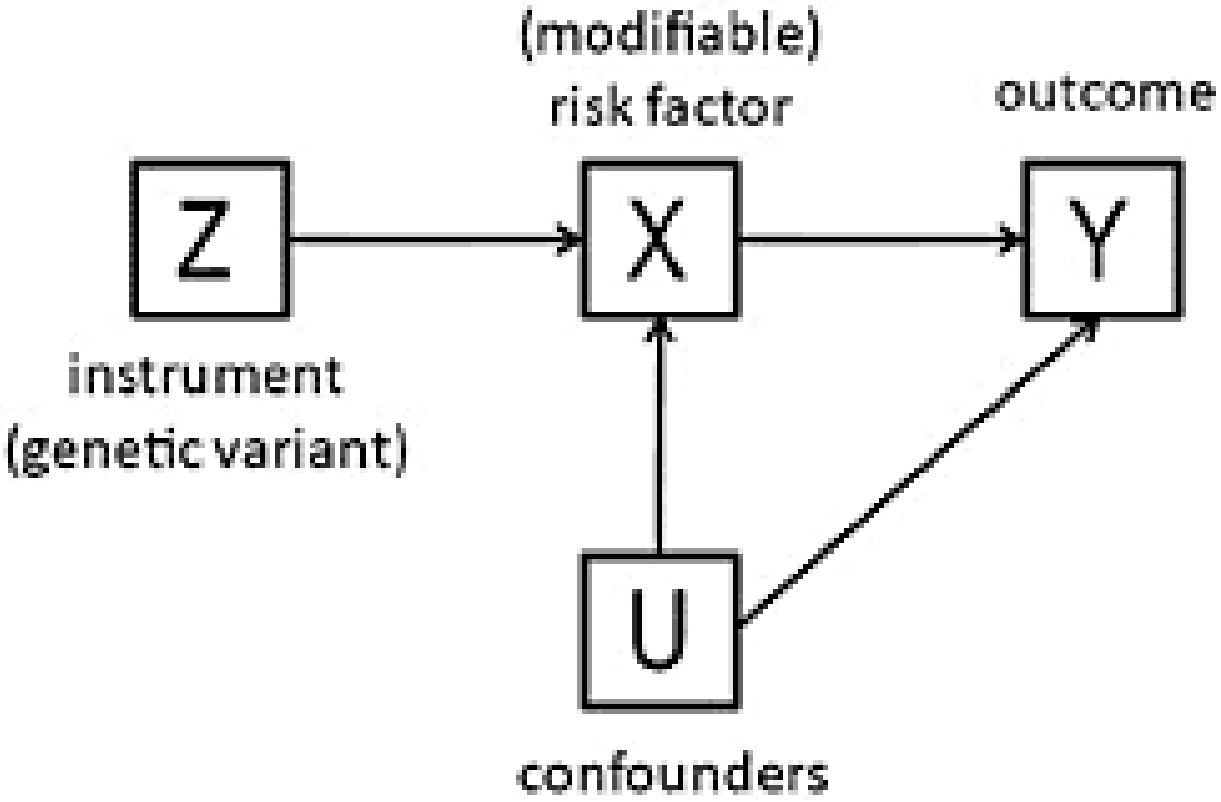
- Have you started to think about causality in your PhD project?
- Is it obvious?
- Share reflections

Randomization without intervening





Genes are used as instrument variable



International Journal of Epidemiology, 2014, 1473–1483
doi: 10.1093/ije/dyu119
Advance Access Publication Date: 6 June 2014
Original article

Original article

High tobacco consumption is causally associated with increased all-cause mortality in a general population sample of 55 568 individuals, but not with short telomeres: a Mendelian randomization study

Line Rode, Stig E Bojesen, Maren Weischer and Børge G Nordestgaard*

Department of Clinical Biochemistry and The Copenhagen General Population Study, Copenhagen University Hospital, Herlev, Denmark and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

*Corresponding author. Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-2730 Herlev, Denmark. E-mail: boerge.nordestgaard@regionh.dk

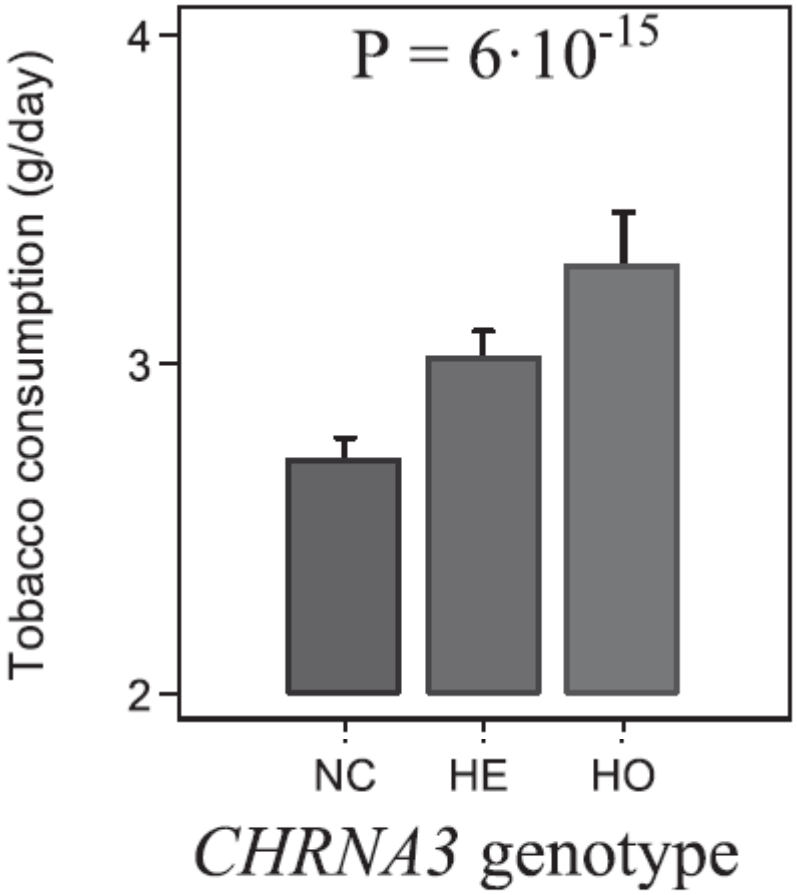
Accepted 19 May 2014

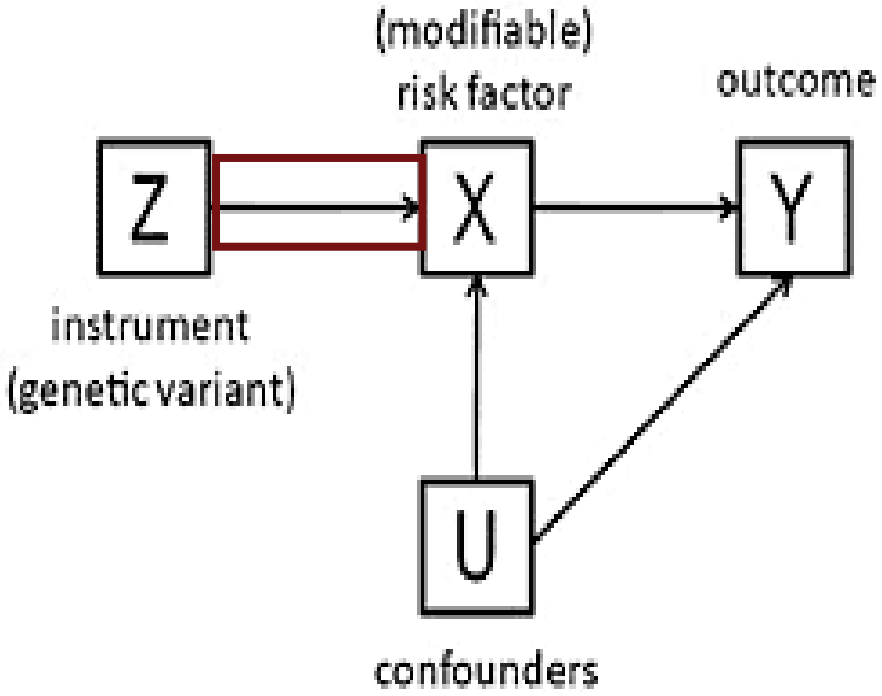
Abstract

Background: High cumulative tobacco consumption is associated with short telomeres and with increased all-cause mortality. We tested the hypothesis that high tobacco consumption is causally associated with short telomeres and with increased all-cause mortality.

Methods: We studied 55 568 individuals including 32 823 ever smokers from the Danish general population, of whom 3430 died during 10 years of follow-up. All had telomere length measured, detailed information on smoking history, and *CHRNA3* rs1051730 genotype, which is associated with tobacco consumption, determined. In a Mendelian

CHRNA3 is associated with smoking





Confounders according to smoking and genotype

Table 1. Baseline characteristics of the 55 568 individuals from the general population according to cumulative tobacco consumption

Characteristic		Ever smokers by quartiles of cumulative tobacco consumption						
	Never smokers (*N*=22 745)	1st quartile (*N*=8205)	2nd quartile (*N*=8206)	3rd quartile (*N*=8206)	4th quartile (*N*=8206)	*P*-trend	*P*-trend telomere length[a]	*P*-trend genotype[b]**
Pack-years, median (IQR)	0	3 (1–5)	12 (9–15)	24 (20–28)	44 (38–56)			
Age, median (IQR)	55 (45–65)	54 (45–65)	55 (46–65)	58 (49–67)	63 (56–71)	<0.001	<0.001	0.20
Male gender, *n* (%)	9039 (40)	2838 (35)	3719 (45)	4031 (49)	5239 (64)	<0.001	<0.001	0.98
Body mass index (kg/m^2), median (IQR)	25.0 (23.0–28.4)	25.0 (22.8–27.6)	25.6 (23.3–28.4)	26.0 (23.5–28.7)	26.6 (24.0–29.5)	<0.001	<0.001	0.09
Alcohol consumption (units per week), median (IQR)	6 (2–12)	7 (3–14)	8 (3–15)	9 (3–17)	11 (4–21)	<0.001	<0.001	0.77
Less than 10 years in school, *n* (%)	5481 (24)	1664 (20)	2148 (26)	2847 (35)	4108 (50)	<0.001	<0.001	0.53
Leisure time physically inactive, *n* (%)	10 883 (48)	3896 (47)	4177 (51)	4623 (56)	4897 (60)	<0.001	<0.001	0.08
Diabetes, *n* (%)	780 (3.4)	224 (2.7)	261 (3.2)	397 (4.8)	633 (7.7)	<0.001	<0.001	0.05
C-reactive protein (mg/l), median (IQR)	1.4 (1.1–2.3)	1.4 (1.1–2.2)	1.5 (1.1–2.4)	1.7 (1.2–2.9)	2.1 (1.3–3.9)	<0.001	<0.001	0.01(ns)

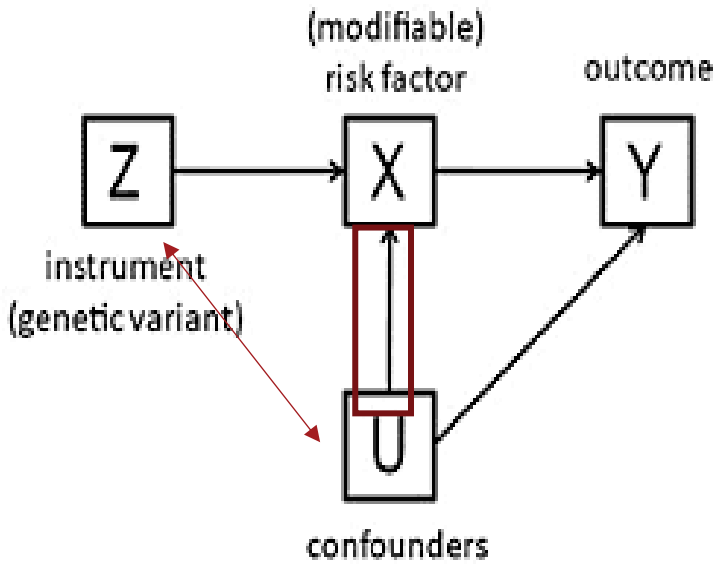
1 unit alcohol ~12 g.

P-trend: Cuzick's extension of the Wilcoxon rank-sum test.

IQR, interquartile range; ns, non-significant after Bonferroni adjustment ($P < 0.05/8 = 0.006$ is considered statistically significa

[a]For details, please see Supplementary Table S1, available as Supplementary data at *IJE* online.

[b]For details, please see Supplementary Table S2, available as Supplementary data at *IJE* online.



Smoking and all-cause mortality

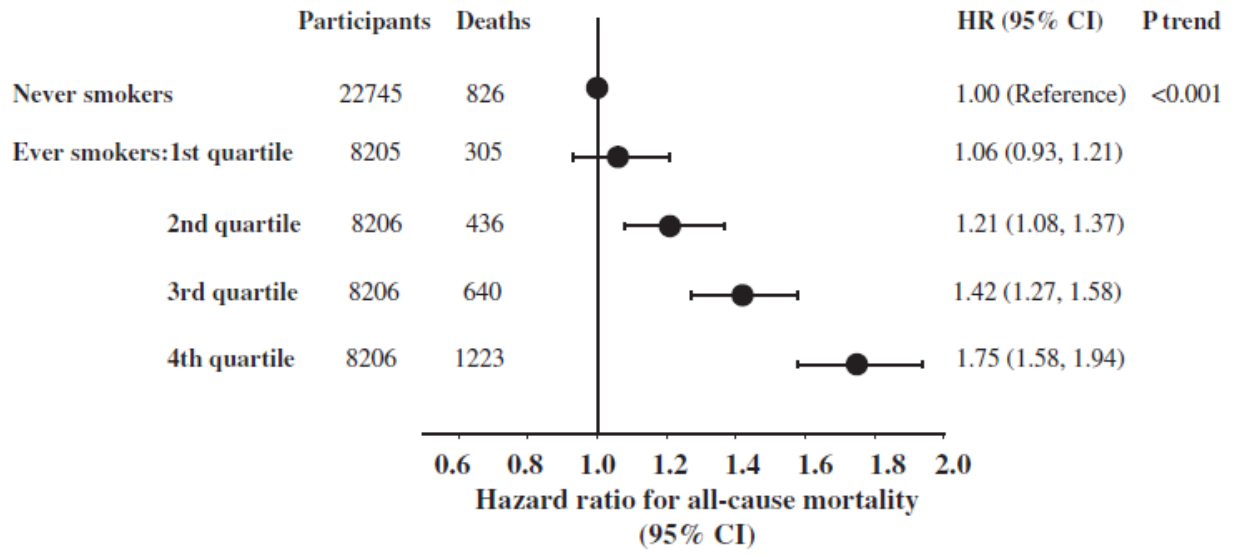
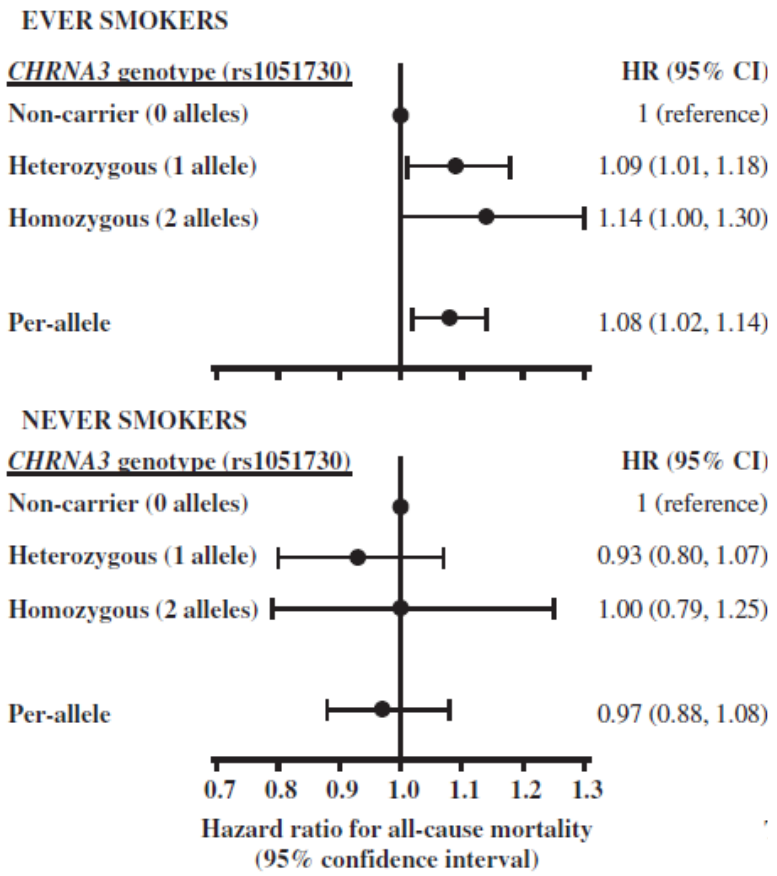


Figure 3. Multivariable-adjusted hazard ratios of all-cause mortality in ever smokers according to quartiles of cumulative tobacco consumption in comparison with never smokers, based on 55 568 individuals from the general population.

Genotype and all-cause mortality



In conclusion, high tobacco consumption is causally associated with increased all-cause mortality. High cumu-

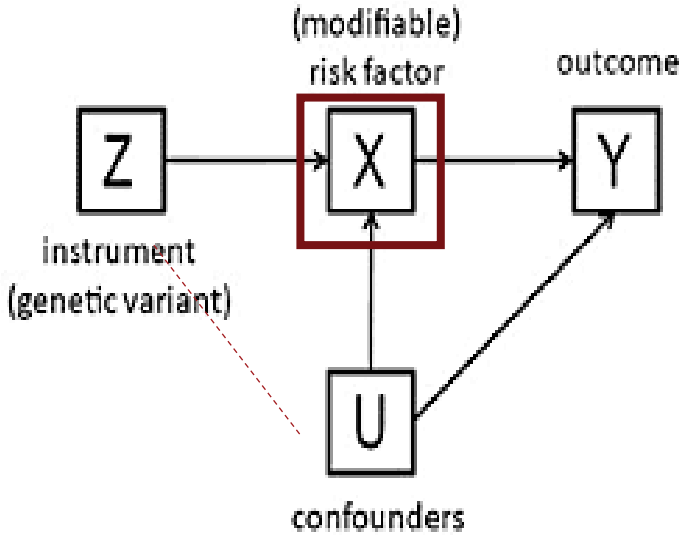


Figure 4. Hazard ratios of all-cause mortality and mean telomere length according to *CHRNA3* genotype in 32 823 ever smokers and 22 745 never smokers from the general population.

Mendelian randomization analyses

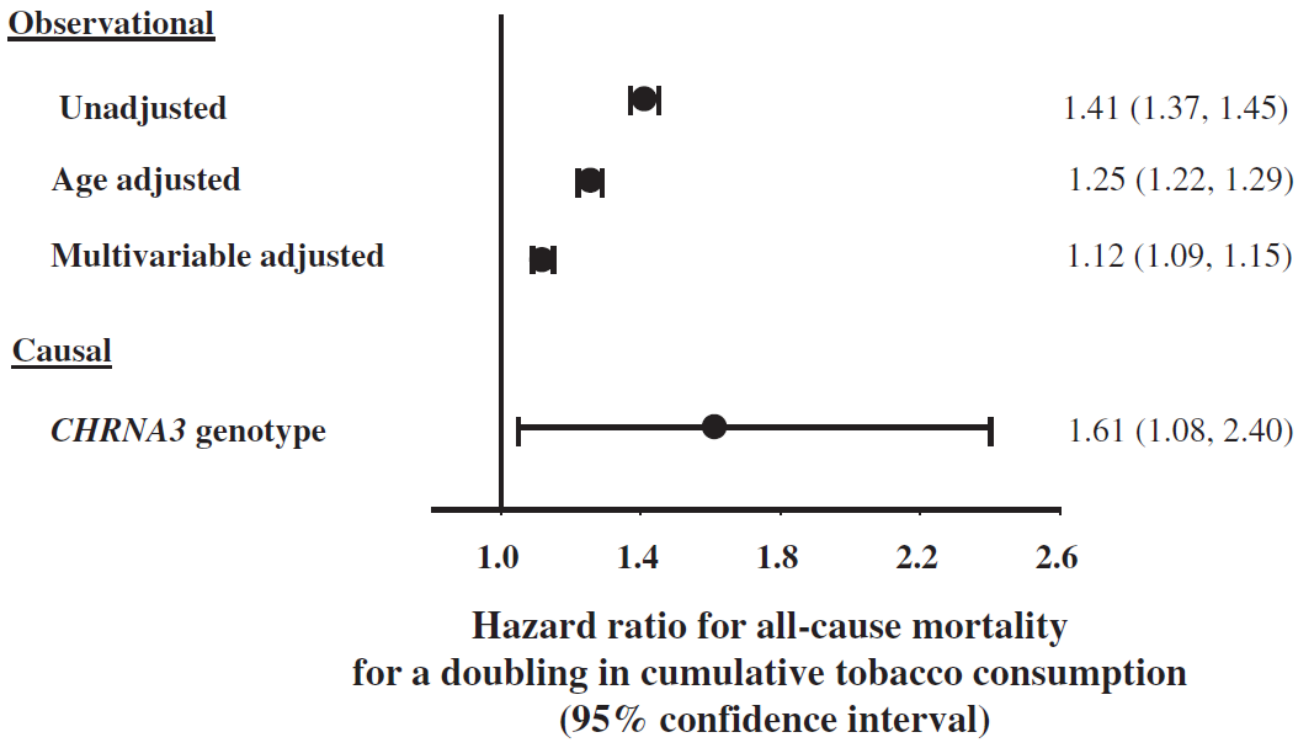
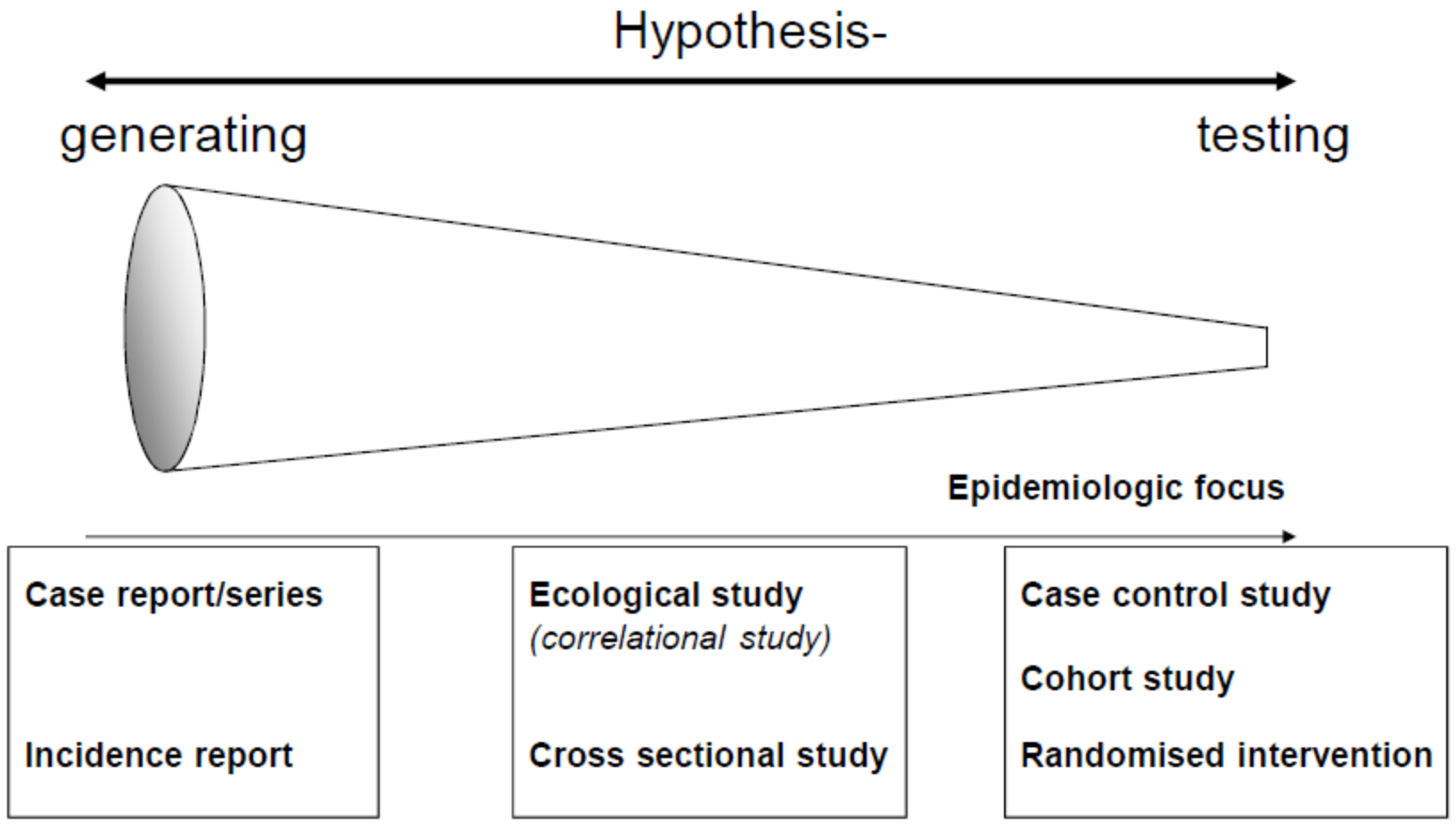


Figure 5. Observational and causal risk of all-cause mortality per doubling in cumulative tobacco consumption based on 32 823 ever smokers from the general population.

Assembling and systemizing evidence



American Journal of Epidemiology

DOI: 10.1093/aje/kwv156

Commentary

Epidemiology and the Tobacco Epidemic: How Research on Tobacco and Health Shaped Epidemiology

Jonathan M. Samet*

* Correspondence to Dr. Jonathan M. Samet, Department of Preventive Medicine, Keck School of Medicine of USC, USC Institute for Global Health, University of Southern California, Soto Street Building, 2001 N. Soto Street, Suite 330A, MC 9239, Los Angeles, CA 90089-9239 (e-mail: jsamet@med.usc.edu).

Initially submitted May 14, 2015; accepted for publication June 10, 2015.

In this article, I provide a perspective on the tobacco epidemic and epidemiology, describing the impact of the tobacco-caused disease epidemic on the field of epidemiology. Although there is an enormous body of epidemiologic evidence on the associations of smoking with health, little systematic attention has been given to how decades of research have affected epidemiology and its practice. I address the many advances that resulted from epidemiologic research on smoking and health, such as demonstration of the utility of observational designs and important parameters (the odds ratio and the population attributable risk), guidelines for causal inference, and systematic review approaches. I also cover unintended and adverse consequences for the field, including the strategy of doubt creation and the recruitment of epidemiologists by the tobacco industry to serve its mission. The paradigm of evidence-based action for addressing noncommunicable diseases began with the need to address the epidemic of tobacco-caused disease, an imperative for action documented by epidemiologic research.

causal inference; epidemiologic methods; smoking; tobacco control

Abbreviation: SHS, secondhand smoke.

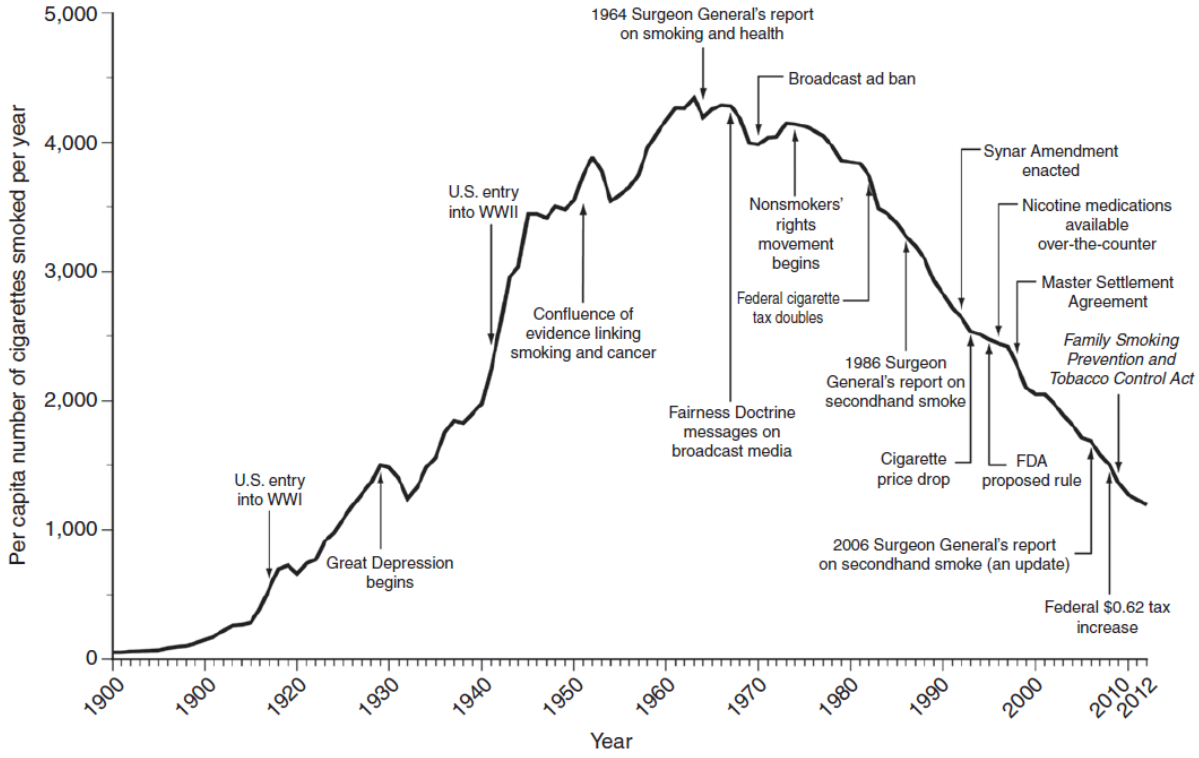


Figure 1. Per capita cigarette consumption and major smoking and health events in adults (≥18 years of age as reported annually by the US Bureau of the Census), United States, 1900–2012. FDA, US Food and Drug Administration; WWI, World War I; WWII, World War II. Reprinted from the Department of Health and Human Services (3), with permission from the Government Publishing Office.

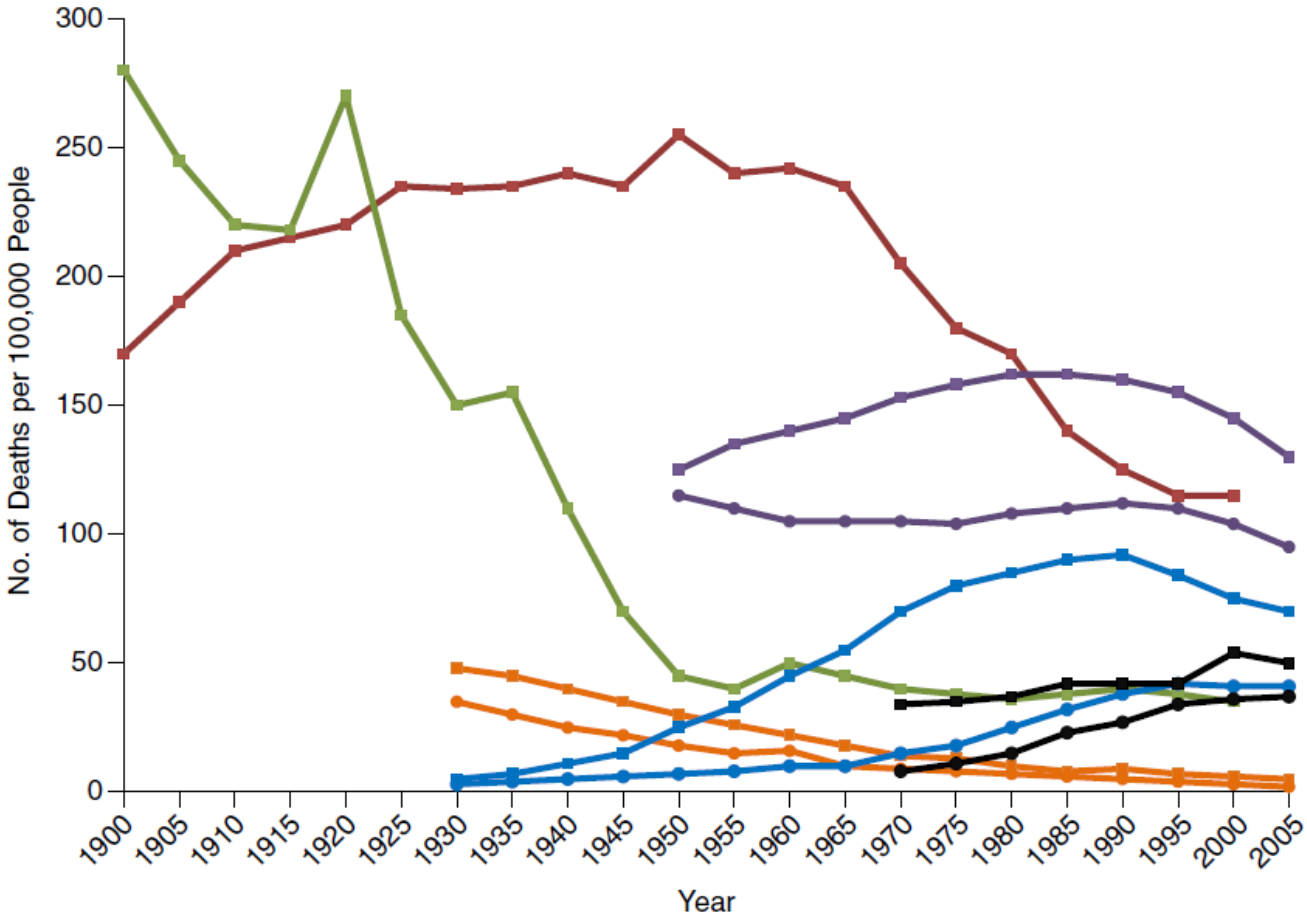


Figure 2. Selected age-adjusted mortality rates in the United States, 1900–2005. Rates are shown for infectious diseases (green), cardiovascular diseases (red), all cancers (purple), lung and bronchus cancer (blue), stomach cancer (orange), and chronic obstructive pulmonary disease (black), with squares representing rates for men and circles representing rates for women. Data on infectious diseases and cardiovascular diseases are from Cutler et al. (48). Data on stomach, lung, and bronchus cancers are from the American Cancer Society (49). Data on all cancers are from the World Health Organization (50).

Exercise

- Would you define the study you are working on now, as descriptive vs. analytical?
- Easy to tell?
- Share reflections